Submitted to *Molecular Biology and Evolution* as an Article
Best fits in the Discoveries section

Why selection might be stronger when populations are small: intron size and density predicts within and between-species usage of exonic splice associated *cis*-motifs.

XianMing Wu[1] and Laurence D. Hurst[*,1]

[1]Department of Biology and Biochemistry, University of Bath, Bath, Somerset, United Kingdom

[*]Corresponding author: E-mail: bssldh@bath.ac.uk.

Abstract

The nearly-neutral theory predicts that small effective population size provides the conditions for weakened selection. This is postulated to explain why our genome is more "bloated" than that of, for example, yeast, ours having large introns and large intergene spacer. If a bloated genome is also an error prone genome might it, however, be the case that selection for error-mitigating properties is stronger in our genome? We examine this notion using splicing as an exemplar, not least because large introns can predispose to noisy splicing. We thus ask whether, owing to genomic decay, selection for splice error-control mechanisms is stronger, not weaker, in species with large introns and small populations. In humans much information defining splice sites is in *cis* exonic motifs, most notably exonic splice enhancers (ESEs). These act as splice-error control elements. Here then we ask whether within and between species intron size is a predictor of the commonality of exonic *cis* splicing motifs. We show that, as predicted, the proportion of synonymous sites that are ESE-associated and under selection in humans is weakly positively correlated with the size of the flanking intron. In a phylogenetically-controlled framework, we observe, also as expected, that mean intron size is both predicted by $N_e.\mu$ and is a good predictor of *cis* motif usage across species, this usage co-evolving with splice site definition. Unexpectedly, however, across taxa intron density is a better predictor of *cis*-motif usage than intron size. We propose that selection for splice-related motifs is driven by a need to avoid decoy splice sites that will be more common in genes with many and large introns. That intron number and density predict ESE usage within human genes is consistent with this, as is the finding of intragenic heterogeneity in ESE density. As intronic content and splice site usage across species is also well predicted by $N_e.\mu$, the result also suggests an unusual circumstance in which selection (for *cis* modifiers of splicing) might be stronger when population sizes are smaller, as here splicing is noisier, resulting in a greater need to control error-prone splicing.

Key words: synonymous mutation, exonic splice enhancer, purifying selection, intron density.

Introduction

Classical nearly-neutral theory proposes that selection will be less efficient as the effective population size (N_e) goes down (Ohta 1973; Ohta 1992; Ohta 1996). In this context, we can for example, interpret the finding that humans have a more "bloated" genome than seen in a species such as yeast which has a large effective population size and a correspondingly "lithe" genome (Lynch and Conery 2003). A lithe genome is one with short intergene spacer, relatively little repetitive sequence, few introns with the few found being relatively small. Might it, however, be the case that, as genomes decay owing to reduced N_e, the error rates of critical processes go up (c.f. Frank 2007)? This might include increased mistranscription, mistranslation, misplicing, incorrect protein folding, incorrect phosphorylation, incorrect subcellular localization *etc.* (Lynch 2007). Might this in turn then result in otherwise paradoxical stronger selection on error mitigation phenotypes when populations are small? Were this so, this would add a novel dimension to the nearly neutral hypothesis, as it would suggest that selection can sometimes be stronger when effective populations sizes are small, because, in this instance, the error rates are higher.

In the paper we examine this possibility by considering splicing error as an exemplar. In particular, we assume 1) that intron sizes tend to increase as N_e declines and that this is largely attributable to genome bloating (Lynch and Conery 2003) and 2) that, within a genome exons flanked by larger introns have noisier splicing. As a consequence, we hypothesise that selection to reduce splice error rates will be more common in species with large introns, typically those with low N_e. Put differently, might humans have gradually expanded their introns through multiple small insertions, each being unable to be resisted by purifying selection, but in the process increased selection on modifiers of splicing in a ratchet-like process (c.f. Frank 2007). The selection to reduce splice error rates we suggest will be manifested, in part, as a higher density of exonic *cis*-modifiers of splicing.

The two assumptions of our hypothesis appear to be reasonable, although the first of these has proven controversial. From phylogenetically uncontrolled correlation based analysis Lynch and

Conery (2003) noted that across a wide span of species, as $N_e.\mu$ declines introns tend to get larger and more common (higher density). $N_e.\mu$ note is the product of effective population size (N_e) and the mutation rate (μ), the single statistic being estimated from population heterozygozity data. The trend in intron size Lynch and Conery attribute to weakening selection as N_e declines, i.e. species with low N_e are less able to eliminate, via purifying selection, weakly deleterious insertion mutations when they occur in introns (and intergenic sequence). This study has, however, been criticised for failing to allow for phylogenic non-independence between data points (Whitney and Garland 2010). Indeed, it was argued that the key result is not robust to proper phylogenetic control (Whitney and Garland 2010). As this $N_e.\mu$ intron size/number correlation is a central tenet of the nearly-neutral interpretation of genome anatomy, we return to this issue employing a phylogenetically controlled mode of analysis and more up to date estimates of $N_e.\mu$, employing both more data and multiple modes of estimation. We show that with these updated estimates, in a phylogenetically controlled framework, $N_e.\mu$ does indeed predict intron dimensions as Lynch and Conery (2003) postulated. We also show, however, that Whitney and Garland had an important objection, as we do not robustly recover this result using the original Lynch and Conery estimates of $N_e.\mu$.

Our second supposition, that larger introns pose a threat to accurate splicing, has received experimental and comparative support. Notably it is observed that experimental insertion of sequence into introns can reduce splice rates (Klinz and Gallwitz 1985; Luehrsen and Walbot 1992; Fox-Walsh et al. 2005; Sironen et al. 2006) and the exons hardest to splice consistently are those flanked by large introns (Bell et al. 1998; Fox-Walsh et al. 2005). Exons flanked by short introns, also associated with high expression levels, tend, by contrast to be subject to less noisy splicing (Pickrell et al. 2010). In the longer term, exons flanked by long introns tend to be those most commonly lost (Kandul and Noor 2009), consistent with splice error rates being too high to sustain the exon. Exactly why exons flanked by larger introns are harder to splice is not fully understood, but one can speculate that if an intron is large, the splice site is harder to locate and the possibility for cryptic splice sites contained within the intron would be higher. The true splice sites need the reinforcement afforded by SR proteins binding to ESEs.

Our hypothesis that selection to reduce splice error rates will be manifested in part as a higher density of exonic *cis*-modifiers of splicing, is, in part, predicated upon the knowledge that *cis*-modifiers of splicing are known to be important in humans. For our genes only approximately 50% of the information defining splice sites is at the splice site, the rest being in *cis* motifs (Lim and Burge 2001). Possibly the most important of these motifs are exonic splice enhancers (ESEs) (Blencowe 2000). The importance of ESEs is well demonstrated by the influence they have on selection on synonymous mutations (Carlini and Genut 2006; Parmley et al. 2006; Cáceres and Hurst 2013). Recent estimates suggest that around 4-5% of synonymous mutations in humans are under purifying selection because they disrupt ESEs (Cáceres and Hurst 2013). Our hypothesis might also predict that this figure might be a little lower in mice than in humans, as humans have on average larger introns. This has yet to be established, but suggestively, while standard nearly-neutral N_e based arguments would more obviously have predicted that selection on synonymous sites should be less common in humans than in rodents (Sharp et al. 1995; Keightley et al. 2005), the reverse seems to be true: an estimated 20% of synonymous mutations under selection in humans but only 10% in mice (Eory et al. 2010).

As *prima facie* support for the notion that selection for splice-error proofing can be more intense when populations are small, we note that inferred centrality of ESEs to splicing in humans contrasts with species, such as yeast, with few/small introns and large populations. *Saccharomyces cerevisiae*, for example, appears not to employ ESEs to reinforce splicing (Spingola et al. 1999; Warnecke et al. 2008). More generally, the modes of selection on synonymous mutations in yeast and mammals appear to be rather different. While in yeast there is easily identified translational selection (whereby codon usage evolves in accord with the tRNA pool), most acute in highly expressed genes (Ikemura 1982, 1985; Kanaya et al. 2001), the same is not robustly found in mammals (Bernardi et al. 1985; Sharp et al. 1995; Kanaya et al. 2001; Duret 2002). Rather, in mammals, selection on synonymous mutations is predominantly at exonic ends where ESEs aggregate (Carlini and Genut 2006; Parmley et al. 2006; Parmley et al. 2007; Cáceres and Hurst 2013). In addition, however, there is evidence for selection on synonymous mutations in mammals mediated by miRNA pairing (Hurst 2006; Brest et al. 2011; Gartner et al. 2013), cotranslational folding (Lawrie et al. 2013) and mRNA structure modulation (Chamary and Hurst 2005; Nackley et al. 2006; Bartoszewski et al. 2010).

Our hypothesis makes a series of intra- and inter-specific predictions. We expect, for example, that within a genome selection on ESEs might be more common in exons neighbouring larger introns. Prior evidence supports the possibility that intron size is an important predictor ESE density, at least within the human genome, ESEs being at a higher density at exon ends in proximity to longer introns (Dewey et al. 2006; Cáceres and Hurst 2013). It is not, however, known if the higher density also implies more ESEs under selection. More generally, it isn't known if all putative ESE sites are functional. The apparent excess near long introns may, for example, reflect simple biased nucleotide content covarying with intron size (Duret et al. 1995). Here then we first ask whether selection on ESE-related synonymous sites might be more common in the vicinity of large introns, controlling for nucleotide usage. To this end we estimate the absolute number of ESE related synonymous sites in proximity to an exon-intron junction that are under selection, as a function of the size of the flanking intron.

Our hypothesis also predicts that ESE usage should vary greatly between species, being greater when populations are small and introns large. Prior evidence suggests that there is indeed considerable between-species variation in exonic *cis*-motif usage. While ESEs are only well described in a handful of species, trends in *k*-mer usage across species in the vicinity of exon ends can be employed as a surrogate measure (Warnecke et al. 2008; Wu et al. 2013). Many *k*-mers are either enriched or depleted in the vicinity of exon junctions, trends in amino acid and codon usage in the vicinity of exon ends being a case in point. These trends are typically well predicted by underlying nucleotide content of the *k*-mers and the extent to which such nucleotides are employed in ESEs (Parmley and Hurst 2007; Cáceres and Hurst 2013), these being commonly purine-rich (Cáceres and Hurst 2013). Indeed, even in a species as distant from humans as *Ectocarpus* (a brown algae), 6-mer trends accord well with known human-described ESEs (Wu et al. 2013). Moreover, species lacking such distortion in *k*-mer usage also tend to be those that do not employ SR proteins to aid splicing, SR proteins being the binding partners of ESEs (Warnecke et al. 2008; Wu et al. 2013). Conversely, trends in *k*-mer usage in the vicinity of exon ends have been employed to define novel splice related exonic motifs (Lim et al. 2011).

Taking the degree of distortion on *k*-mer usage in the vicinity of exon ends as a metric of the extent of *cis*-motif usage for splice control, prior studies report considerable variation between taxa in the number of *k*-mers affected (Warnecke et al. 2008; Wu et al. 2013). Here then we ask whether we can account for this variation in terms of between-species variation in the size of introns and the effective population size. For compatibility with prior studies we employ in frame 3-mers, i.e. codons. Prior evidence suggests that *cis*-motif usage, measured this way, may be most prevalent in species with more intronic sequence (Warnecke et al. 2008), but whether it is intron size or number that matters is not clear. Also suggestive of a relationship between ESE usage and intron dimensions, we recently showed that *Ectocarpus*, a species very distant from mammals and unusual in also having large introns, has extensive *cis*-motif usage, these motifs corresponding to ESEs well described in humans (Wu et al. 2013).

These prior analyses have, however, been confounded by a difference between species in the number of exons sampled and by not controlling for phylogeny. They also don't distinguish between intron density and intron size as predictors, whilst our model relates to intron size. We here ask in a phylogenetically explicit framework a) whether mean intron size is a predictor of a species usage of *cis* motifs and b) whether it is a stronger predictor than intron density (the number of introns per bp of CDS). To date we are unaware of experimental evidence suggesting that intron density should predict ESE usage. This being so, if the selection across taxa for ESEs is mediated by changes in intron size alone, then intron density should not be a good predictor. In addition we employ a compound predictor, this being the ratio of CDS size to gene size that factors both intron density and mean intron size. If only intron size is relevant then this compound predictor should be no better a predictor than mean intron size. Finally, we can ask how such trends in *cis* motif usage correlate with N_e or rather $N_e.\mu$, this metric estimated from intraspecific polymorphism levels. Given prior evidence that ESE usage and the nucleotides defining the splice site co-evolve (Fairbrother et al. 2002; Dewey et al. 2006), we also address splice site usage as a function of $N_e.\mu$ and intron size.

Results

1. Selection on synonymous mutations is more common when the flanking intron is large

The question as to whether selection on *cis*-splice motifs is more commonplace when the flanking intron is larger has two components: first, to what extent are such motifs under purifying selection as a function of the size of the flanking intron and second, how common are such motifs as a function of the size of the flanking intron.

Human ESEs are slower evolving when the flanking intron is larger, but this is likely to be mutational.

Are ESEs slower evolving than non-ESE sequence towards exon ends and is both the rate of evolution and the degree of constraint modulated by the size of the flanking intron? To address this we consider human-macaque aligned sequence and classified exon ends in terms of the size of the flanking intron. As the exon ends are so small, to minimize estimation noise we consider for each intron size range the concatenation of all exon end alignments so as to provide a single estimate of Ks for each intron size class. We compare the synonymous rate of evolution in and out of ESE sequence.

To consider whether a hexamer might be an ESE motif we took advantage of a recent analysis which derived two sets of motifs that were agreed on by the majority of ESE discovery analyses as being ESEs (Cáceres and Hurst 2013) and hence provide gold standard datasets with low false positive rates. As these are datasets that are intersects of independent data sets, in which at least three of four putative ESE datasets agree that a given hexamer motif is an ESE motif, we follow that prior nomenclature and refer to these as INT3 and INT3_400. Of the four original input data sets one (Ke et al. 2011) presented a liberally defined set of ESEs and a more conservative top 400 set. As these two input sets are non-independent, the prior authors (Cáceres and Hurst 2013) built two intersect data sets: one where the liberal set was employed and one in which the more conservative 400 strong data set was employed. The two resulting three way intersect sets were thus termed INT3 (84 hexameric motifs with the liberal set employed) and INT3_400 (54 hexameric motifs with the top 400 hexamers employed) respectively.

As can be seen (Fig.1.A, Fig.1.B) the ESE sequence evolves slower at synonymous sites than does non-ESE, as previously shown (Parmley et al. 2006; Cáceres and Hurst 2013). The difference between ESE and non-ESE may be a consequence of differences in mutation rate owing to skewed nucleotide content of ESEs. To examine this we simulated sets of randomized pseudoESE sets that are the same size and drawn from the same underlying nucleotide content as the true ESE sets. We then match these pseudoESEs against the sequence alignments to determine the rate of evolution associated with these. These sets evolve faster than the true ESEs suggesting that ESEs are indeed (as commonly reported) under purifying selection (Fig 1), even allowing for biased nucleotide content.

More striking, we observe an evident negative correlation between Ks of ESEs at exon ends and the size of the flanking intron (INT3_400: 5' rho = -0.50, P = 0.027; 3' rho = -0.64, P = 0.003; INT3: 5' rho = -0.53, P = 0.018; 3' rho = -0.74, P = 3×10^{-4}). This is consistent with stronger purifying selection on *cis*-splicing motifs or mutation rate differences covarying with intron size. That we see a commensurate decrease in Ks of the "non-ESE" sequence as a function of intron size might reflect either a) purifying selection in exon ends is generally stronger in the vicinity of large introns, possibly because the definition of non-ESE is too liberal and includes much sequence that is functional splice related motif or b) the mutation rate in exons is lower in the vicinity of larger introns. To examine the latter possibility we compare Ks of exon cores as a function of the size of neighbouring introns (we consider the size of the 5' and 3' intron separately), under the presumption that little or no sequence in exon cores will modulate splicing. We observe that Ks of cores also show a decreasing tendency as intron sizes increases (Fig 1). While we can conclude that the reduced rate of evolution of ESEs, compared to nonESE and pseudoESE in the same exons, is not solely mutational in origin, we cannot then exclude the possibility that the low rate of synonymous evolution at exon ends in the vicinity of large intron is at least in part owing to genomically regional mutation rate biases in the vicinity of large introns.

Consideration of the rate of synonymous evolution of exon cores also permits us to define the approximate degree of constraint operating on ESE at exon ends as:

Flank ESE constraint = [Ks core – Ks ESE flank] / Ks core.

This may be conservative, but it is noteworthy that Ks non-ESE flank, Ks pseudoESE and Ks core are all approximately of the same magnitude (Fig.1.A, Fig.1.B), much higher than Ks ESE flank. In the absence of purifying selection on ESE at exon flanks, in excess of that at exonic cores, the degree of constraint should be zero. We observe that the level of constraint, thus defined, operating on ESEs at exon flanks is not significantly related to the size of the flanking intron, although Spearman's rho is positive in all incidences (Fig.2.A, Fig.2.B; supplementary table 1.1). What can reasonably be concluded is that selection on ESEs is not obviously weaker in the vicinity of large introns. To estimate the number of sites under selection at exon flanks, we need in addition to factor in not just the level of constraint, but also ESE density. This we consider next.

Allowing for increased ESE density in proximity to large introns, selection on synonymous sites associated with ESEs is (slightly) more common when introns are larger.

It has previously been reported that ESE density tends to be a little higher in the vicinity of longer introns (Dewey et al. 2006; Cáceres and Hurst 2013). We replicate this by partial correlation analysis between ESE density and three intronic dimensions (supplementary table 2.1). For ESE dataset INT3, both 5' and 3' show significant correlation between ESE density and mean intron size (5' rho = 0.03, $P = 9 \times 10^{-4}$; 3' rho = 0.03, $P = 9 \times 10^{-4}$). However for the smaller INT3_400 ESE data set, 3' correlation is not significant (INT3_400: 5' rho = 0.06, $P = 2 \times 10^{-13}$; 3' rho = -0.01, P = 0.14). More marginal results at exonic 3' ends is a common theme in our analyses which we comment on later.

To evaluate the net effect of flanking intron size (constraint and increased density) we calculate the proportion of synonymous sites under ESE-related constraint at exon flanks as: flank ESE constraint × ESE density. It is no surprise that the net effect of flanking intron size on proportion of sites under selection is an increasing function, albeit only weakly so, as both underlying trends are positive. However, using the conservative binning method (N=20) the trend is not significant. This may well reflect a limited sample size (N=20). To avoid this problem we instead calculate

the regression line of logarithm value of flank intron size versus ESE constraint (using unbinned data). Using this regression line we then estimate the mean ESE constraint for exon flanks given the size of the neighbour intron. For each exon individually we then calculate ESE density × regression estimated constraint. We find in all cases a positive and highly significant Spearman's rank correlation (INT3_400: 5' rho = 0.439, P = 0; 3' rho = 0.123, P = 2.7×10^{-31}; INT3: 5' rho = 0.070, P = 2.5×10^{-13}; 3' rho = 0.646, P = 0).

Some of these latter values appear unusually high, which may relate to our interpolation method which smoothes out noise resulting from the tiny number of sites contributing to constraint estimates at individual exon ends. Moreover, this method doesn't allow for potential covariance with intron number and intron density. To examine this we analyse gene level (rather than individual exon end) metrics. For each gene we consider the intron density, intron number and mean intron size. In addition we consider the constraint revealed in concatenated exon flanks and concatenated exon cores. A small minority of genes have no synonymous site evolution in exon flank ESEs giving a constraint of unity. As Spearman's correlation isn't necessarily robust to tied values we thus also replicated analyses using Goodman Kruskall gamma test with P determined by simulation (Supplementary table 3.1). To minimize estimation noise, we require for all genes a minimum of 102 bp of concatenated sequence. We then ask whether mean intron size is related to constraint on ESE at flanks.

We find that mean intron size is positively and significantly correlated with 5' but not 3' ESE constraint (Supplementary table 3.1). This trend is weak (rho 0.048-0.056) but is reported using both statistics and both ESE data sets. Intron density and intron number are not significant predictors. Partial spearman correlation also reports mean intron size as a significant predictor (Supplementary table 3.2). To some degree these results are not robust to increasing the minimum threshold length for analysis from 102 bp to 150 bp or higher. This, however, appears to be a consequence of reduced sample size. We resampled genes by the number of 150 bp cutoff group from the gene pool of 102 cutoff group and repeated 1000 times to find how often the intron dimensions can significantly predict the constraints. For mean intron size, only in a little over than 60% of resamplings can we still see significant Spearman partial correlation with 5' ends ESE constraint for both datasets. For Goodman and Kruskal's gamma, the commensurate

figure is around 42% (Supplementary table 3.3). This accords with the trends being weak and hence sensitive to sample size reductions.

We conclude that in the human genome mutations in ESEs at exon ends are probably more commonly under selection when the flanking intron is larger, the effect being mostly mediated by an increased ESE density. This result in turn suggests that disease-associated mutations might be slightly more common in exon ends in the vicinity of large introns, but the effect appears to be modest.

We note that as regards this result we are agnostic as to the cause. This may be a direct effect of intron size or owing to a covariance between intron size and splice site strength, possibly with expression level as a covariate. Our intention here is not to distinguish between these explanations, but simply to suppose that this evidence provides *prima facie* support to the hypothesis that an increase in intronic dimensions within a species can be coupled with more selection for *cis*-modifiers of splicing. We note, however, that a model that supposes that ESEs are used more in exons next to long introns might be a means to increase elongation rate, to compensate for the time to process the longer intron, is not well supported (see Supplementary Table 4.1~ 4.2).

2. The ratio of mature CDS to gene size is the best predictor of between-species *cis*-motif usage.

Given the above result and prior experimental and comparative data on the difficulty of splicing exons when the neighbouring intron is large (see Introduction), we might expect that mean intron size would be a predictor of the commonality of the usage of exon flank *cis*-modifiers of splicing. To establish the latter we consider, for 30 highly phylogenetically dispersed species, the proportion of codons or amino acids that show significant trends in their usage as a function of the distance from an exon-intron junction, these metrics having been shown previously to correspond well with ESE motif usage (Parmley and Hurst 2007; Parmley et al. 2007; Warnecke et al. 2008; Cáceres and Hurst 2013). Using a Bayesian comparative framework, we can then ask whether mean intron size is indeed a good predictor of *cis*-motif usage. We find that it is (Table

1). While our measure of the degree of this trend controlling for the size of exonic input data set is the preferred metric, we show that usage of an uncontrolled metric (employing all valid exons within a species) does not distort the picture (Table 1). Hereafter we employ the sample size controlled metric exclusively, unless mentioned otherwise.

As a control test, we consider intron density. Intron density, measured as number of introns per kb of mature CDS, holds no information regarding the size of the introns and hence if size is the key variable density should be irrelevant. Unexpectedly, not only do we find that density is a predictor of the extent of *cis*-motif usage, we also observe that it is consistently a better predictor than intron size (the BayesTrait score is higher in all modes of analysis: Table 1). Given this surprising result we ask whether a metric that considers the net effect of density and size might be an even better predictor. To this end we employ the ratio of mature CDS to gene size (alias immature transcript size). This is consistently the best predictor (Table 1). We conclude that intron size alone is not adequate to describe the between-species trends in *cis*-motif usage and that density effects are also of relevance. The logic of the importance of the density effects we discuss below.

3. Evidence for coevolution of splice site and *cis*-motif usage

Prior evidence suggests that ESE usage is higher in proximity to certain splice sites (Berget 1995; Graveley 2000). One possibility is that "weak" splice sites might be more in need of the reinforcement offered by flanking ESEs (Fairbrother et al. 2002). In support of this ESE density appears to be stronger in proximity to "weak" splice sites (Dewey et al. 2006; Plass et al. 2008; Cáceres and Hurst 2013). To ask whether ESE usage across species was predicted by relative usage of different splice sites, we investigated all splice sites across 30 species. The splice sites we represented as four-letter nucleotide strings, nucleotides of exons in upper case, nucleotides of introns in lower case. After phylogenetic correction, BayesTraits provided very strong evidence for correlation between usage of *cis*-motif and usage of two splice sites ("AGgt" and "agGT") (Table 2). This indicates a preference of exonic splice associated *cis*-motifs to these specific splice sites. These results indicate that the trends in *cis*-motif usage across species reflect in part co-evolution with splice site usage.

4. **$N_e.\mu$ predicts intronic dimensions**

Given that intronic dimensions predict *cis* motif usage across taxa, what, we can ask, predicts intronic dimensions across taxa? An attractive proposal is that introns and intronic sequence accumulate owing to weakened selection against insertions associated with reduced N_e. Previously, Lynch and Conery (2003) have argued, in a phylogenetically uncontrolled analysis, that intronic size can be well understood in the context of such a nearly-neutral model. They posit that as N_e reduces so selection becomes weaker and the ability of a species to resist weakly deleterious insertions (both new introns and new sequence within extant introns) is in turn reduced. Thus they predict large introns and high density of introns in species with low N_e.

Their analysis has been criticized on numerous fronts, not least of which is the assumption of $N_e.\mu$ is a good predictor of the behaviour of N_e alone (Daubin and Moran 2004) (a problem our analysis is also sensitive to). Further, they estimated $N_e.\mu$ for a sample of species often employing limited sequence data. Perhaps most importantly, their analysis was criticised for failing to control for phylogenetic structure, in effect assuming a star phylogeny (Whitney and Garland 2010). This same follow-up analysis, employing a phylogenetically explicit method failed to observe a relationship between genome size parameters and N_e. We return to this issue employing three methods to estimate $N_e.\mu$, three metrics of intronic content and a fully controlled phylogenetic methodology.

Three $N_e.\mu$ values of this study show very significant correlations between themselves; however, our estimates of $N_e.\mu$ do not correlate well with those of Lynch and Conery (Table 3, Supplementary Fig. S1, The blue line indicates the SMA regression). We find that our $N_e.\mu$ estimates robustly predict all three intronic dimensions in the expected direction (Table 4). By contrast, we can replicate Whitney and Garland's failure to detect such a correspondence: after phylogenetic correction, although there is a strong evidence to support the correlation between $N_e.\mu$ values of Lynch and Conery's study and the ratio of mature CDS to gene size, these $N_e.\mu$ values do not correlate well with intron density and mean intron size (Table 4). We suggest that

the paucity of data contributing to the Lynch and Conery estimates of $N_e.\mu$ is the major issue with their analysis.

5. $N_e.\mu$ predicts splice site usage but not *cis* motif usage

The above sets of results suggest a simple narrative to explain *cis*-motif usage across species. As N_e declines, so introns become more abundant and larger, owing to the weakening of purifying selection (result 4 above). A consequence of this is that small insertions may accumulate in a ratchet-like manner. Similarly, splice sites might decay. Both splice site decay and the increase in intron size cause increases in the rate of mis-splicing compensated by increased usage of exonic *cis*-motifs. Within genomes, the argument goes, this is reflected in a higher density of functional *cis*-motifs in the flanks of exons that neighbour large introns (result 1) and associated with particular splice sites (Cáceres and Hurst 2013). Thus, selection on synonymous mutations at exon flanks is more common when the flanking intron is large (result 1 above) and species with on average larger introns have more *cis*-modifiers (result 2), these being especially common when certain splice sites become more common (result 3). Additionally, consistent with ESE-splice site coevolution, we see intra-specifically that AGgt exons are flanked by larger introns (supplementary table 5.1), consistent with splice site - ESE - intron size three way coevolution. We would thus expect that $N_e.\mu$ should also in turn predict the usage of *cis*-splice modifiers and splice sites.

The latter result we find to be robustly supported, at least for 5' end splice site usage. More specifically, the correlations between $N_e.\mu$ values and the usage of "AGgt" (i.e. 5' splice site) are very strong, while those about the usage of "agGT" (3' splice site) are weak (Table 5).

Do we also find that $N_e.\mu$ predicts *cis*-motif usage? This result we have yet to demonstrate. The prediction we make is that species with low $N_e.\mu$ will be species with more common skews in codon or amino acid usage owing to selection for *cis*-modifiers of splicing. Unexpectedly, despite having observed all prior correlations ($N_e.\mu$ predicts intron dimensions and splice site usage, intron dimensions and splice site usage predict *cis*-motif usage), we fail to recover a trend whereby *cis*-motif usage is predicted by $N_e.\mu$ (Table 6). For the $N_e.\mu$ estimator *S* there may be a

weak trend but for others there is no evidence. Employing the sample size uncorrected measure of the number of trends removes any weak trend reported for *S* (Table 6). We conclude that we find evidence that splice site usage, but not *cis*-motif usage, correlates with $N_e.\mu$.

6. Alternative splicing rate does not explain *cis*-motif usage

One reason that $N_e.\mu$ might not predict *cis*-motif usage is that other covariates are important and mask any effect. A potentially key covariable might be the frequency of alternative splicing. We observed previously that the brown algae *Ectocarpus* has a striking number of codons and amino acids showing skews in usage in the vicinity of exon junctions, many more indeed than humans (Wu et al. 2013). This we hypothesised may reflect the low rate of alternative splicing that we could detect. If alternative splicing is rare in a species, then more of the annotated exons will be under selection to be properly spliced more of the time. Alternatively, ESEs might modulate alternative splicing, which is more common in "complex" species (Chen et al. 2014), typically with low N_e. Note that these two models make opposite predictions.

To provide an assessment of this we consider transcript depth controlled estimates of the rate of alternative splicing for 14 species (Chen et al. 2014). We find strong evidence to support the correlation between alternative splicing rates with the ratio of mature CDS to gene size. While intron density is a better predictor of *cis* motifs than is intron size, the correlation between alternative splicing rates and mean intron size is better than that with intron density (Table 7). Between-species differences in alternative splicing rates do not, however, predict between-species trends in *cis*-motif usage very well (Supplementary table 5.2). We conclude that while alternative splicing rates and intronic dimensions covary, the former appears not to explain trends in *cis*-motif usage.

7. Is the commonality of decoy splice sites the main driver of splice associated *cis*-motif usage?

Why might it be that the best between-species predictor of *cis*-motif usage was not simply mean intron size, but an aggregate measure of size and density? From the logic that we laid out

(difficulty of exon junction recognition in the context a large flanking intron), this is perhaps unexpected. We suggest the problem may be one of decoy splice sites. Imagine a gene with one large intron and no residues elsewhere downstream of the true 3' splice site that might be recognized as a possible acceptor site. Would such a gene have error-prone splicing? We would suggest not if there is a unique strong site (the true acceptor site) compatible with splicing. By contrast, by definition, the same gene with two introns must have at least two putative acceptor sites. Thus the more introns and the weaker the splice sites, the more potential there is for mis-splicing.

This suggests then a simple explanation for why intron density matters. We assume that SR proteins bound to the immature RNA accumulate at exon ends bound to ESEs. A given 5' splice site, we assume also, tends to attach to the perceived nearest 3' splice site, this being identified by the accumulation of ESEs and SR proteins. The extent of accumulation of ESEs we suggest is a function of the chance the splice site might be "missed". Strong splice sites in close proximity (short introns) are unlikely to be missed and hence need little reinforcement. By contrast ESEs are needed more in the vicinity of larger introns as the ability to find the nearest 3' splice site is harder owing to the distance and because the number of decoy sites is higher, i.e. when the density of introns is higher. However, whether it is density *per se* or absolute number of introns that is key is not immediately transparent, as it is unclear whether the absolute proximity of decoy splice sites to the "real" splice site is relevant. If physical proximity is relevant then density may matter, if not absolute number may be more important.

Such a model makes an intragenomic prediction, namely that controlling for intron length, intron density or number should predict ESE density. From the partial correlation analysis between ESE density and three intronic dimensions (mean intron size, intron density, and intron number), all 5' end calculations show very significant partial correlations, regardless of the choice of ESE data set. At the 3' end the result is less clear. For INT3 ESE data set, the correlation between ESE density and intron density is not significant, while intron number and intron size are predictors. At 3' ends all partial correlations for INT3_400 are not significant (supplementary table 2.1). The correlation with intron number is perhaps the most revealing, suggesting that density *per se* functions as a proxy to absolute number and hence that exon size considerations are not so

relevant. Further, these results suggests that, while ESE usage at 5' and 3' ends of exons is usually considered to be symmetrical in humans (motifs commonly found at 5' ends tend to be common at 3' ends (Warnecke et al. 2008; Lim et al. 2011)), that at least as regards intron density mediated effects 5' and 3' ends are under different modes of selection. The suggestive evidence that net selection on ESEs is better correlated with intron length for 5' ends than 3' ends supports the same proposition, as does the 5'- 3' difference in splice site predicted by $N_e.\mu$.

If the problem faced is one in which downstream exons and introns presenting decoy splice sites, then we might also expect a difference in ESE density within a gene, as different exons have a different number of downstream introns and exons and hence a different number of potential decoy splice sites. We address this by comparing the ESE density at the 5' end of the second exon in a gene and the 5' ESE density at the last but one exon in genes with at least four exons. We do not employ the very last exon owing to possible constraints on nucleotide content in the vicinity of the stop codon.

We find strong evidence that intragene location matters, with ESE density higher earlier in a gene. From comparing the ESE density at the 5' end of the second exon in a gene and the 5' ESE density at the last but one exon in genes with at least four exons, the medians of ESE density of last but one exons and second exons are about 2 fold different (INT3 ESE density: 0.0909 and 0.1739, INT3_400 ESE density: 0.0882 and 0.1739), in last but one and second exon respectively (supplementary table 6.1). To examine the significance of this we perform a paired test, comparing the ESE density within the same gene between the two exon 5' flanks. Results are as expected of the decoy splice site model. For INT3 dataset, the number of genes which show ESE density of the second exon to be higher than that of last but one exon, reaches 491 and the number where ESE density of second exon is relatively lower is 381 (binomial test $P= 2.6 \times 10^{-5}$). For INT3_400 dataset, the corresponding values are 332 (ESE density of second exon is higher) and 240 (ESE density of second exon is lower), again supporting a higher density in second exons (binomial test $P= 2.0 \times 10^{-5}$).

To test whether the trend is owing to confounding effects of proximal intron size, we employed Mann Whitney U test to analyse residuals of a loess regression while 5' proximal intron size is

being controlled. Results are again as predicted by the decoy model. We again find a significantly higher ESE density at 5' end of second exons compared with last but one exons (Mann Whitney U test comparing residuals of 5' intron size versus ESE density, INT3: $P = 2.42\times10^{-71}$, INT3_400: $P = 5.32\times10^{-46}$). A within-gene paired test on residuals from above loess regression supports the same conclusions (INT3_400: number of higher second exon residuals = 386, number of lower second exon residuals = 289, binomial test $P = 2.85 \times 10^{-5}$; INT3: number of higher second exon residuals = 566, number of lower second exon residuals = 404, $P = 3.25 \times 10^{-8}$). A higher density of ESEs earlier in a gene is, we suggest, consistent with the decoy model given that early exons by definition have more downstream splice sites than do later ones. It also suggests a novel (to our knowledge) model of splicing reinforcement that is different in different sections of the same gene.

Discussion

We conjectured that reduced N_e might lead to larger introns and weakened splice sites, which in turn could lead to stronger selection for motifs that keep in check the increase in the degree of error-prone splicing. All results bar one support this. We find synonymous sites are more commonly under selection within humans when exons are flanked by larger introns (largely because more sites function as *cis* motifs), that intronic dimensions and splice site usage predict *cis*-motif usage across species and that $N_e.\mu$ predicts intronic dimensions and splice site usage (we note that this tidies up the prior objection that in a phylogenetic framework the results of Lynch and Conery don't hold (Whitney and Garland 2010)). In addition, we find that intra-specifically, exons flanked by large introns both have higher ESE density and greater usage of AGgt, consistent with coevolution between splice site, ESEs and intron size. What we don't observe is that $N_e.\mu$ predicts *cis*-motif usage.

Given the support for the hypothesis from all but one of the tests, we suggest it would be premature to reject the hypothesis out of hand. Indeed, one possibility is that our estimation of $N_e.\mu$ is either too rough or otherwise flawed. It is striking, for example, that our estimation and that of Lynch and Conery do not correlate well, despite being based on the same underlying

premise. Moreover there might be a systematic issue with all polymorphism based attempts to estimate $N_e.\mu$, this being that the expected correlation between N_e and heterozygozity appears to be much weaker than predicted by the neutral model (which forms the basis for $N_e.\mu$ estimation). Gillespie (2001) argues that the approximate invariance (or weak positive correlation) between N_e and heterozygozity is owing to an increased rate of positive selection when populations are large, thereby causing regular collapses of heterozygozity owing to hitchhiking type effects. We do not wish to comment on the veracity of this claim, but simply wish to note that of all the variables that we have employed, $N_e.\mu$ is the one we have least confidence in, both as regards its estimation and its interpretation. Recent evidence that intra-specific diversity is predicted by life history traits (Romiguier et al. 2014) adds to the notion that a relationship between $N_e.\mu$, deduced from heterozygozity data, and the strength of selection may be compounded by covariates. Nonetheless, we observe that $N_e.\mu$ robustly predicts intronic dimensions and splice site usage, suggesting that it is perhaps not too poor an estimator.

Whilst we have framed the above hypotheses and results in the context of the nearly-neutral model, the same results might, however, also be consistent with a model in which increasing *cis*-motif usage across taxa reflects greater tissue or cell type diversity, ESEs then operating as providers of tissue-specific alternative splice patterns. It is indeed observed that species with more cell types do have more alternative transcripts (Chen et al. 2014). Might this coupling be explained by increased usage of ESEs? Our and other results suggest not. We observe no relationship between *cis*-motif usage and alternative splicing rates. Moreover, there is no strong prior evidence to suppose that ESE usage is a modulator of alternative splicing. Indeed, while our intersect data sets find no difference in ESE density between alternative and constitutive exons (Cáceres and Hurst 2013), an experimentally defined set of exonic splice modifiers (Ke et al. 2011) found a much higher ESE density in constitutive than in alternative exons. Earlier reports also indicated that, while conserved alternative exons have very low rates of evolution, this was not owing to especially strong constraint on ESEs (Parmley et al. 2006; Cáceres and Hurst 2013). These results thus suggest that ESEs are not there as elements to control alternative splicing forms, but rather to make more robust the splicing of constitutive exons, especially those with weak splice sites. For these reasons we suggest that higher transcript diversity in species with

small population sizes/multiple tissue types is not an easily defendable explanation for the trends in *cis*-motif usage.

An unexpected result was that in the between-species comparison, intron size is by no means the best intron-dimension predictor of *cis*-motif usage. Rather a combination of size and density is a much better predictor. We propose a decoy splice spite model as a potential explanation. This model correctly predicts intragenomic and intragenic trends, highlighting the selection on the earliest exons as being especially acute. The intragenic trend may however have an alternative explanation, namely that it is simply more damaging to missplice an early exon than it is to missplice a later exon. For example, the downstream effects of a frame-shifting splice event may be different for the two. It is not so obvious that such an argument can explain the intragenomic, intergenic trends (i.e. mean intron size, intron density, and intron number all independently predict 5' ESE usage). This model, and the apparent asymmetry between 5' and 3' effects are, we suggest, worthy of further scrutiny.

Materials and Methods

Exon and intron sequences from 30 species

From "Table Browser" of UCSC (http://genome.ucsc.edu/cgi-bin/hgTables, last accessed January 23, 2014) and FTP site of NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes, last accessed January 23, 2014), we obtained all available genes from 30 species (*A.carolinensis, A.gambiae, A.thaliana, B.distachyon, C.elegans, C.jacchus, C.neoformans, D.discoideum, D.melanogaster, D.rerio, E.siliculosus, G.gallus, G.gorilla, H.sapiens, I.tridecemlineatus, M.gallopavo, M.mulatta, M.musculus, O.latipes, O.sativa, P.abelii, P.falciparum, P.tetraurelia, P.troglodytes, S.cerevisiae, S.pombe, S.purpuratus, S.scrofa, T.rubripes, X.tropicalis*). Sequences without normal start codon (ATG) and stop codons (TAA, TAG, and TGA), the genes have internal stop codons, ambiguous nucleotides ("N"), and those did not contain introns were all taking away from the dataset for further analysis (Supplementary table 7.1).

Determining trends in amino acid and codon usage

In previous analyses, codons preferred near exon ends were well predicted by the composition of

experimentally defined ESEs (Parmley and Hurst 2007; Cáceres and Hurst 2013). We thus presume that the frequency of distorted codon or amino acid usage in vicinity of exon junctions is a fair measure of *cis* splice motif usage. The trend in usage of each codon and amino acid was investigated as a function of the distance from the exon–intron boundary up to a distance of 34 codons (to accord with an earlier analysis (Warnecke et al. 2008)). The 5' and 3' ends were analyzed separately with the codon in direct proximity to the boundary being eliminated and the first and last exons being excluded. For each codon and amino acid under consideration, we determined, after Bonferonni correction, rho and P value by 2-tailed Spearman correlation of proportional usage as a function of distance from the boundary. A negative rho, indicates a codon or amino acid that is preferred near exon ends, whereas a positive value implies a codon or amino acid preferred at exonic cores and avoided at the ends. For each species we then calculate the proportion of codons or amino acids showing significant skew both at 5' and 3' ends across all exons and consider this the metric of *cis* motif usage for that species.

In order to ensure that these trends comparisons are not affected by the different number of exons in different species, for each species, we made a pool of exons and abstracted 5000 exons from it randomly with replacement (For each repeat of 30 species, 30 data sets were established with each containing 5000 exons). After 100 repetitions of this sampling process, we obtained the mean usage trends of amino acids and codons for each species by the same method mentioned above. We counted up the number of amino acids or codons that showed a significant rho score in the sample size controlled subsampling and employed this as our metric of the extent of *cis* motif usage (Supplementary table 7.2). We also report results for a sample size uncorrected metric.

Splice-related genomic traits

Based on the data sets of genes saved, we calculated three parameters: X (mean CDS length/gene length), N (introns per kb exon) and M (mean intron size) for each species (Supplementary table 7.3).

Phylogenetic tree with branch length

A text file containing an ID list of the 30 species was uploaded to "Taxonomy Browser" of NCBI (http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi, last accessed January 23, 2014), and then we saved the Taxonomy Common Tree, which has no branch length, in PHYLIP format. To obtain the branch lengths of the phylogenetic tree, a multiple sequence alignment was needed. We searched for candidate orthologs through the orthologous database OrthoDB (http://cegg.unige.ch/orthodb7, last accessed January 23, 2014) and HomoloGene (http://www.ncbi.nlm.nih.gov/homologene, last accessed January 23, 2014) of NCBI, by taking gene function (related with temperature, air pressure, oxygen concentration, or acid-base properties) into consideration. 10 orthologous genes (Supplementary table 7.4), conserved in Eukaryotes, were finally used to make alignments by M-Coffee (http://tcoffee.crg.cat/apps/tcoffee/do:mcoffee, last accessed January 23, 2014) separately. All the 10 alignments were merged into one. Gblocks (Castresana 2000; Talavera and Castresana 2007) was employed to eliminate/minimize poorly aligned positions and divergent regions (Supplementary material S1) and converted from Newick format into Nexus within R (parameters used can be found in Supplementary material S2).

Through the RelTime application (Kumar et al. 2012; Tamura et al. 2012; Tamura et al. 2013), a phylogenetic tree with branch lengths was constructed by loading the taxonomy common tree and the alignment into MEGA (version 6). We tested the correlation of results from two models (Jones-Taylor-Thornton model and WAG model) when selecting "Gamma Distributed" of "Rates and Patterns" and other default parameters. There is a very strong correlation between the branch length estimates from the two models (Spearman correlation: $\rho=0.9970$, $P=4.17\times10^{-65}$; Supplementary Fig. S2). We regarded the mean of the results as the final branch lengths (Supplementary Fig. S3).

Correlation between amino acid/codon usage trends and the genomic traits after phylogenetic correction

The application "Continuous" of BayesTraits (Pagel 1999) was used to study correlations between amino acid/codon usage trends and the genomic traits by Markov chain Monte Carlo (MCMC) method. According to the suggestion from the manual of BayesTraints, we abstracted the last harmonic mean from the result file, and took it as an estimation of marginal likelihood, to

calculate the "Log BF" value and further test whether there is strong evidence for the correlation after phylogenetic correction.

Correlation analysis of $N_e.\mu$ values in phylogenetic manner

To calculate $N_e.\mu$ values of the species, several R packages (ape (Paradis et al. 2004), PopGenome (Pfeifer et al. 2014), adegenet (Jombart 2008; Jombart and Ahmed 2011), pegas (Paradis 2010), Geiger (Harmon et al. 2008)) and DnaSP (Rozas and Rozas 1995; Librado and Rozas 2009) were used to analyze the published allelic sequences from "PopSet" (http://www.ncbi.nlm.nih.gov/popset, last accessed January 23, 2014) and "Nucleotide" (http://www.ncbi.nlm.nih.gov/nuccore, last accessed January 23, 2014) database of NCBI. Intron sequences were considered as candidates first and, if there is no intron sequence, CDS sequences were chosen for the analysis in which only synonymous sites number, as the segregating sites number, were input the program (Supplementary table 8.1). Finally, $N_e.\mu$ values ($N_e.\mu$ for per site from Pi, $N_e.\mu$ for per site from S and $N_e.\mu$ for per site from Eta) of each species were obtained for further correlation analysis (Supplementary table 8.2). $N_e.\mu$ values from this study were compared, by Spearman's correlation and SMA regression of R package "lmodel2", with the $N_e.\mu$ values (Lynch and Conery 2003) published previously (Supplementary table 8.3).

By using BayesTraits, in phylogenetic manner, we did correlation analysis of $N_e.\mu$ values (both the values from our study and from Lynch and Conery's study) with amino acid/codon usage trends and three intronic dimensions (mean CDS length/gene length, intron density and mean intron size) (Supplementary table 8.4).

Comparison of selection on synonymous mutations with different flanking intron size

A list of human macaque orthologs was obtained from ENSEMBL (Flicek et al. 2014). Only those defined as 1:1 orthologs were employed. The respective genes were extracted from human CDS build GRCh37.74 and Macaque build MMUL_1.74. These were aligned using MUSCLE 3.8.31 (Edgar 2004) at the protein level, the nucleotide alignment being built from the protein alignment using a custom script (AA2NUC). Exon and intron sizes for the relevant human genes were obtained via ENSEMBL. Any gene whose CDS length did not match that specified in the BioMart (Kasprzyk 2011) derived annotation file was excluded. The alignment of the exons was

derived from the exon dimensions specified (naturally with allowance for indels). Only internal (not first or last exons) exons from the macaque-human comparison were employed.

We considered only exons longer than 2 × 69 bp and considered the 5' 69 bp as the 5' end and 3' 69 bp as the 3' end. The alignments were masked with two consensus ESE candidate data sets, INT3 and INT3-400, these being intersect datasets between four high coverage databases of putative ESE sequences (Cáceres and Hurst 2013). One of the datasets presents a large sample of putative ESEs and a second (N=400) top hit sample. As these are non-independent the intersect data sets either employ the full sample (INT3) or the reduced sample (INT3-400). We could thus, employing these two separately, define sites that were ESE and sites that were possibly not ESE (although as these two sets were conservative, there are likely to be true ESEs in the non-ESE class of sequence). For both ESE and non-ESE masking of the alignments, we then concatenate all exon ends as a function of twenty different flanking intron sizes, thereby making estimation of Ks less noisy. We also compared Ks of exon cores (69 bp of core region in each exon) as a function of the size of neighbouring introns after concatenating core sequences in each bin. For each of the sets of concatenated exon ends and cores, both ESE and non-ESE, we estimate Ks using PAML (version: PAML 4.7, Default parameters are used, codon model = 2)(Yang 2007).

To exclude the possibility that any trends seen are not artifacts of skewed nucleotide content between ESE and non-ESE sequence, we generated pseudo ESE sets containing the same number of random hexamers with, on average, the same nucleotide content as each ESE set. Then the same test as above was performed 100 times repeatedly for each pseudo ESE set. The average value with standard error bar from these nucleotide controls is displayed in the plots (Fig.1.A, Fig.1.B).

Partial correlation between ESE density and three intronic dimensions

ESE density and three intronic dimensions (mean intron size, intron density, and intron number) were obtained using custom perl scripts. When two of the intronic dimensions are controlled, partial correlation between ESE density and another intronic dimension was analysed by R program, pcor.test (Kim and Soojin 2006) (Supplementary table 2.1).

Correlation between flank ESE constraint and three intronic dimensions

Based on the alignment dataset of human macaque orthologs, Ks core and Ks ESE flank (both 5' 69 bp and 3' 69 bp, exons are shorter than 138 bp were all regarded as flank region), of each gene, were calculated after concatenating all flank and core sequences. We set up three criteria for concatenating sequence to select genes for correlation analysis (1. ESE flank > 102 bp, Core region > 102 bp; 2. ESE flank > 150 bp, Core region > 150 bp; 3. ESE flank > 201 bp, Core region > 201 bp). Then three intronic dimensions (mean intron size, intron density, and intron number) and Flank ESE constraint of each gene were obtained by our perl script. For INT3 and INT3_400 datasets, we explored the correlation between Flank ESE constraint (defined in results) and three intronic dimensions, considering 5' and 3' ends of exons separately, by partial Spearman's correlation (Supplementary table 3.2).

We evaluate the net effect of flanking intron size (constraint and increased density) by calculating the proportion of synonymous sites under ESE-related constraint at exon flanks as: flank ESE constraint × ESE density. Instead of using the conservative binning method (N=20), we calculated the regression line of logarithm value of flank intron size versus ESE constraint and for each exon individually calculate ESE density × constraint, where constraint is estimated via interpolation of this regression line, given the intron size. Then we examine whether the Spearman's rank correlation between ESE density × constraint and the logarithm value of intron size is significant.

Furthermore, Goodman and Kruskal's gamma, by using program "rcorr.cens" from R package "Hmisc" (http://biostat.mc.vanderbilt.edu/Hmisc, https://github.com/harrelfe/Hmisc) was been carried out in above analysis to avoid affects of tied observations. P value, which shows whether Goodman and Kruskal's gamma is significant, comes from p=(n+1)/(m+1) where n is the number of gamma values calculated after randomly shuffled the variables representing flank ESE constraint and meanwhile greater than the observed gamma and m is 1000, this being the number of times of shuffling (Supplementary table 3.1).

To make sure the above result is not affected by sample size artefacts, we did a resampling test by abstracting genes by the number in the 150 bp cutoff group from the gene pool in the 120 cutoff

group and repeated the two types of correlation analysis 1000 times. We report the proportion of random subsamplings that still provide a significant correlation prior to multitest correction.

Comparison of ESE density between second exons and last but one exons

In the human gene dataset, we selected from genes with four or more exons, second exons and last but one exons, which are all greater than 138 bp. We calculated 5' exon end ESE density and 5' flank intron size of these two exon categories in each gene. To control for the effect of flank intron size we analysed the residuals from loess regression of 5' end ESE density predicted by 5' intron size (supplementary table 6.1). Both analyses were repeated for the two ESE data sets (INT3 and INT3_400). Significance was assayed via a binomial test counting the absolute number of genes having a higher density at the 2nd exon than the last but one, versus the opposite. If ESE density was no different these were ignored.

Relationship between transcriptional elongation rate and ESE density

We used publicly available data from a genome-wide elongation rate study (Veloso et al. 2014) to investigate the relationship of ESE density with transcriptional elongation rate (around 450 genes were selected due to requirement of ESE density calculation, supplementary table 4.1) and also correlated the elongation rates with several genic dimensions used in our study (supplementary table 4.2).

Supplementary Material

Supplementary tables 1–8, figures S1–S3, and other material S1(".htm" file)-S2(".txt" file) are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by a University Research Studentship from the University of Bath to XianMing Wu and a Wolfson Royal Society Research Merit Award to L.D.H. The work was funded in part by Medical Research Grant MR/L007215/1. We thank Mike Lynch for access to data and advice regarding data.

Tables

Table 1. Evidence for phylogenetically-controlled correlation between amino acid/codon usage trends and the genomic traits.

We employ two metrics of skews at exon ends, the number of codons showing a significant skew and the number of amino acids showing a significant skew. For each, in addition we report results wherein for each species all relevant exons are employed and a second metric where the input sample size is the same for all species (5000 randomly chosen exons). In the latter instance we consider the mean number of significant trends from multiple samplings of 5000 randomly chosen exons.

	All exons (AA)	All exons (Codon)	Random 5000 exons (AA)	Random 5000 exons (Codon)
Log BF (Y^b~X^c) [a]	48.241	39.394	31.923	42.027
Log BF (Y~N^d)	37.484	29.202	24.055	32.294
Log BF (Y~M^e)	20.145	15.018	12.214	18.410

[a]Log BF (log Bayes factor) =2*(log [harmonic mean (complex model)]-log [harmonic mean (simple model)]), is the test statistic of BayesTraits which gives the information of evidence for correlated evolution: weak evidence (<2), positive evidence (>2), strong evidence (5-10), very strong evidence (>10). All Log BF values in the table are >10, so the evidence from all correlations is very strong.

[b] Y=Proportion of amino acids/codons showing significant trends

[c] X=mean CDS length/gene length

[d] N=introns per kb exon

[e] M=mean intron size

Table 2. *Cis*-motif usage correlates significantly with usage of "AGgt" and "agGT" splice sites

	All exons (AA)	All exons (codon)	Random 5000 exons (AA)	Random 5000 exons (codon)
Log BF (Y[a]~P1[b])	39.0359	31.7091	26.4632	33.7518
Log BF[d] (Y~P2[c])	52.1594	64.0366	76.8355	58.1153

[a]: Proportion of amino acids/codons showing significant trends,
[b]: Proportion of AGgt (Capital letter: exon, small letter: intron),
[c]: Proportion of agGT,
[d]:Log BF (log Bayes factor) =2*(log [harmonic mean (complex model)]-log [harmonic mean (simple model)])
All Log BF values in the table are >10, so the evidences of all correlations (positive) are very strong.

Table 3. Spearman's Correlation analysis results for $N_e.\mu$ values of this study and the prior study of Lynch and Conery. We compare our three different estimators for $N_e.\mu$, (Eta, Pi and S) and Lynch's single estimate.

	rho	rho^2	*P*
$N_e.\mu$ _Eta ~ $N_e.\mu$ _Lynch	0.093	0.009	0.765
$N_e.\mu$ _Pi~ $N_e.\mu$ _Lynch	0.165	0.027	0.591
$N_e.\mu$ _S~ $N_e.\mu$ _Lynch	0.093	0.009	0.765
$N_e.\mu$ _Eta ~ $N_e.\mu$ _S	0.996	0.991	0.000
$N_e.\mu$ _Pi ~ $N_e.\mu$ _Eta	0.970	0.941	0.000
$N_e.\mu$ _Pi ~ $N_e.\mu$ _S	0.975	0.951	0.000

Table 4. Evidence for phylogenetically-controlled correlation between $N_e.\mu$ values and splice-related Genomic Traits. We employ our three different estimators for $N_e.\mu$, (Eta, Pi and S) and Lynch's single estimate.

	X[b]	N[c]	M[d]
Log BF ($N_e.\mu$ _Pi ~ Splice-related Genomic Traits)[a]	15.762	23.424	41.057
Log BF ($N_e.\mu$ _S ~ Splice-related Genomic Traits)	14.572	22.590	39.944
Log BF ($N_e.\mu$ _Eta ~ Splice-related Genomic Traits)	13.988	22.695	40.367
Log BF ($N_e.\mu$ _Lynch[f] ~ Splice-related Genomic Traits)	5.290	0.989	-0.587

[a]Log BF (log Bayes factor) =2*(log [harmonic mean (complex model)]-log [harmonic mean (simple model)]), is the test statistic of BayesTraits which gives the information of evidence for correlated evolution: weak evidence (<2), positive evidence (>2), strong evidence (5-10), very strong evidence (>10). All Log BF values in the table are >10, so the evidences of all correlations are very strong.
[b] X=mean CDS length/gene length
[c] N=introns per kb exon
[d] M=mean intron size
[f]this $N_e.\mu$ value is from previous study (Lynch and Conery 2003)

Table 5. Evidence for phylogenetically-controlled correlations between $N_e\mu$ values and usage of "AGgt" (very strong) and "agGT" (weak) splice sites using three estimators of $N_e\mu$, namely Pi, S and Eta

	$N_e\mu$_Pi	$N_e\mu$_S	$N_e\mu$_Eta
Log BF ($N_e\mu$[a] ~ P1[b])	22.7225	19.1016	20.6161
Log BF[d] ($N_e\mu$ ~ P2[c])	1.6456	-0.1762	0.6543

[a]: Three types of $N_e\mu$ ($N_e\mu$_Pi, $N_e\mu$_S, $N_e\mu$_Eta),
[b]: Proportion of AGgt (Capital letter: exon, small letter: intron),
[c]: Proportion of agGT,
[d]:Log BF (log Bayes factor) =2*(log [harmonic mean (complex model)]-log [harmonic mean (simple model)])

Table 6. Little evidence for a phylogenetically-controlled correlation between $N_e.\mu$ values and amino acid/codon usage trends (Y). We employ our three different estimators for $N_e.\mu$, (Eta, Pi and S) and four metrics of *k*-mer usage.

	All exons (AA)	All exons (codon)	Random 5000 exons (AA)	Random 5000 exons (codon)
Log BF ($N_e.\mu$ _Pi ~ Y[b])[a]	-0.486	-2.065	-2.693	-4.436
Log BF ($N_e.\mu$ _*S* ~ Y)	-1.383	-0.206	1.514	0.728
Log BF ($N_e.\mu$ _Eta ~ Y)	0.534	-0.520	-2.038	1.079

[a]Log BF (log Bayes factor) =2*(log [harmonic mean (complex model)]-log [harmonic mean (simple model)]), is the test statistic of BayesTraits which gives the information of evidence for correlated evolution: weak evidence (<2), positive evidence (>2), strong evidence (5-10), very strong evidence (>10). All Log BF values in the table are < 2, so the evidences of all correlations are weak.

[b]Proportion of amino acids/codons showing significant trends

Table 7. Evidence for correlation between alternative splicing rates and Splice-related Genomic Traits

	X[d]	N[e]	M[f]
Log BF (ASL1[b] ~ Splice-related Genomic Traits)[a]	5.259	2.782	7.299
Log BF (ASL2[c] ~ Splice-related Genomic Traits)	8.714	4.589	9.500

[a]Log BF (log Bayes factor) =2*(log [harmonic mean (complex model)]-log [harmonic mean (simple model)]), is the test statistic of BayesTraits which gives the information of evidence for correlated evolution: weak evidence (<2), positive evidence (>2), strong evidence (5-10), very strong evidence (>10)
[b]ASL1: Average no. of ASEs per gene (residual of the polynomial regression between num of ESTs [col. O] and ASL [col. U])
[c]ASL2: Average no. of ASEs per gene (residual of the linear regression between the log-transformed num of ESTs [col. O] and ASL [col. U])
[d]mean CDS length/gene length
[e]introns per kb exon
[f]mean intron size

Figure Legends

Fig.1. Rate of synonymous evolution in ESE and nonESE sequence at exon ends as a function of the Log of flanking intron size for two ESE datasets (Fig.1.A: INT3, Fig.1.B: INT3_400). In addition to Ks of ESE and nonESE we also show Ks of exon core domains and pseudoESE, i.e. hexamers of the same underlying nucleotide content as ESEs but not necessarily identified as being functional ESE. We consider 20 intron size bins apportioned so that all bins contain the same number of exon ends for concatenation, the numbers given reflecting the upper intron size limit of each bin.

Fig.2. The degree of selective constraint on ESE sequences at exon ends as a function of the Log of flanking intron size for two ESE datasets (Fig.2.A: INT3, Fig.2.B: INT3_400). For definition of constraint see main text. For intron size definition see Fig 1. Note that in all cases constraint appears stronger when intron sizes are larger, although using 20 bins the trends are not significant.

Reference

Bartoszewski RA, Jablonsky M, Bartoszewska S, Stevenson L, Dai Q, Kappes J, Collawn JF, Bebok Z. 2010. A synonymous single nucleotide polymorphism in DeltaF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J Biol Chem* 285:28741-28748.

Bell MV, Cowper AE, Lefranc MP, Bell JI, Screaton GR. 1998. Influence of intron length on alternative splicing of CD44. *Mol Cell Biol* 18:5930-5941.

Berget SM. 1995. Exon recognition in vertebrate splicing. *J Biol Chem* 270:2411-2414.

Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F. 1985. The mosaic genome of warm-blooded vertebrates. *Science* 228:953-958.

Blencowe BJ. 2000. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* 25:106-110.

Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, Vouret-Craviari V, Mari B, Barbry P, Mosnier JF, Hébuterne X, et al. 2011. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat Genet* 43:242-245.

Cáceres EF, Hurst LD. 2013. The evolution, impact and properties of exonic splice enhancers. *Genome Biol* 14:R143.

Carlini DB, Genut JE. 2006. Synonymous SNPs provide evidence for selective constraint on human exonic splicing enhancers. *J Mol Evol* 62:89-98.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540-552.

Chamary JV, Hurst LD. 2005. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol* 6:R75.

Chen L, Bush SJ, Tovar-Corona JM, Castillo-Morales A, Urrutia AO. 2014. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol Biol Evol* 31:1402-1413.

Daubin V, Moran NA. 2004. Comment on "The origins of genome complexity". *Science* 306:978; author reply 978.

Dewey CN, Rogozin IB, Koonin EV. 2006. Compensatory relationship between splice sites and exonic splicing signals depending on the length of vertebrate introns. *BMC Genomics* 7:311.

Duret L. 2002. Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev* 12:640-649.

Duret L, Mouchiroud D, Gautier C. 1995. Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. *J Mol Evol* 40:308-317.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.

Eory L, Halligan DL, Keightley PD. 2010. Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. *Mol Biol Evol* 27:177-192.

Fairbrother WG, Yeh RF, Sharp PA, Burge CB. 2002. Predictive identification of exonic splicing enhancers in human genes. *Science* 297:1007-1013.

Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. 2014. Ensembl 2014. *Nucleic Acids Res* 42:D749-755.

Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. 2005. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc Natl Acad Sci U S A* 102:16176-16181.

Frank SA. 2007. Maladaptation and the paradox of robustness in evolution. *PLoS One* 2:e1021.

Gartner JJ, Parker SC, Prickett TD, Dutton-Regester K, Stitzel ML, Lin JC, Davis S, Simhadri VL, Jha S, Katagiri N, et al. 2013. Whole-genome sequencing identifies a recurrent functional synonymous mutation in melanoma. *Proc Natl Acad Sci U S A* 110:13481-13486.

Gillespie JH. 2001. Is the population size of a species relevant to its evolution? *Evolution* 55:2161-2169.

Graveley BR. 2000. Sorting out the complexity of SR protein functions. *RNA* 6:1197-1211.

Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24:129-131.

Hurst LD. 2006. Preliminary assessment of the impact of microRNA-mediated regulation on coding sequence evolution in mammals. *J Mol Evol* 63:174-182.

Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 2:13-34.

Ikemura T. 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. *J Mol Biol* 158:573-597.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403-1405.

Jombart T, Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27:3070-3071.

Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T. 2001. Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J Mol Evol* 53:290-298.

Kandul NP, Noor MA. 2009. Large introns in relation to alternative splicing and gene evolution: a case study of Drosophila bruno-3. *BMC Genet* 10:67.
Kasprzyk A. 2011. BioMart: driving a paradigm change in biological data management. *Database (Oxford)* 2011:bar049.
Ke S, Shang S, Kalachikov SM, Morozova I, Yu L, Russo JJ, Ju J, Chasin LA. 2011. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res* 21:1360-1374.
Keightley PD, Lercher MJ, Eyre-Walker A. 2005. Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol* 3:e42.
Kim S-H, Soojin VY. 2006. Correlated asymmetry of sequence and functional divergence between duplicate proteins of Saccharomyces cerevisiae. *Mol Biol Evol* 23:1068-1075.
Klinz FJ, Gallwitz D. 1985. Size and position of intervening sequences are critical for the splicing efficiency of pre-mRNA in the yeast Saccharomyces cerevisiae. *Nucleic Acids Res* 13:3791-3804.
Kumar S, Stecher G, Peterson D, Tamura K. 2012. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics* 28:2685-2686.
Lawrie DS, Messer PW, Hershberg R, Petrov DA. 2013. Strong purifying selection at synonymous sites in D. melanogaster. *PLoS Genet* 9:e1003527.
Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.
Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. 2011. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc Natl Acad Sci U S A* 108:11093-11098.
Lim LP, Burge CB. 2001. A computational analysis of sequence features involved in recognition of short introns. *Proc Natl Acad Sci U S A* 98:11193-11198.
Luehrsen KR, Walbot V. 1992. Insertion of non-intron sequence into maize introns interferes with splicing. *Nucleic acids research* 20:5181-5187.
Lynch M. 2007. The origins of genome architecture. Sunderland, Mass.: Sinauer Associates ; Basingstoke : Palgrave [distributor].
Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302:1401-1404.
Nackley AG, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskyi O, Makarov SS, Maixner W, Diatchenko L. 2006. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* 314:1930-1933.
Ohta. 1992. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst* 23: 263–286.
Ohta T. 1996. The current significance and standing of neutral and neutral theories. *Bioessays* 18:673-677; discussion 683.
Ohta T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature* 246:96-98.
Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877-884.
Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26:419-420.
Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20:289-290.
Parmley JL, Chamary JV, Hurst LD. 2006. Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. *Mol Biol Evol* 23:301-309.
Parmley JL, Hurst LD. 2007. Exonic splicing regulatory elements skew synonymous codon usage near intron-exon boundaries in mammals. *Mol Biol Evol* 24:1600-1603.
Parmley JL, Urrutia AO, Potrzebowski L, Kaessmann H, Hurst LD. 2007. Splicing and the evolution of proteins in mammals. *PLoS Biol* 5:e14.
Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Mol Biol Evol* 31:1929-1936.
Pickrell JK, Pai AA, Gilad Y, Pritchard JK. 2010. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet* 6:e1001236.
Plass M, Agirre E, Reyes D, Camara F, Eyras E. 2008. Co-evolution of the branch site and SR proteins in eukaryotes. *Trends Genet* 24:590-594.
Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*.
Rozas J, Rozas R. 1995. DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Comput Appl Biosci* 11:621-625.
Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF. 1995. DNA sequence evolution: the sounds of silence. *Philos Trans R Soc Lond B Biol Sci* 349:241-247.
Sironen A, Thomsen B, Andersson M, Ahola V, Vilkki J. 2006. An intronic insertion in KPL2 results in aberrant splicing and causes the immotile short-tail sperm defect in the pig. *Proc Natl Acad Sci U S A* 103:5006-5011.
Spingola M, Grate L, Haussler D, Ares M, Jr. 1999. Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *RNA* 5:221-234.
Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56:564-577.
Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. 2012. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci U S A* 109:19333-19338.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30:2725-2729.

Veloso A, Kirkconnell KS, Magnuson B, Biewen B, Paulsen MT, Wilson TE, Ljungman M. 2014. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res* 24:896-905.

Warnecke T, Parmley JL, Hurst LD. 2008. Finding exonic islands in a sea of non-coding sequence: splicing related constraints on protein composition and evolution are common in intron-rich genomes. *Genome Biol* 9:R29.

Whitney KD, Garland T, Jr. 2010. Did genetic drift drive increases in genome complexity? *PLoS Genet* 6.

Wu X, Tronholm A, Cáceres EF, Tovar-Corona JM, Chen L, Urrutia AO, Hurst LD. 2013. Evidence for deep phylogenetic conservation of exonic splice-related constraints: splice-related skews at exonic ends in the brown alga Ectocarpus are common and resemble those seen in humans. *Genome Biol Evol* 5:1731-1745.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586-1591.

Figures

Fig.1. Rate of synonymous evolution in ESE and nonESE sequence at exon ends as a function of the Log of flanking intron size for two ESE datasets (Fig.1.A: INT3, Fig.1.B: INT3_400). In addition to Ks of ESE and nonESE we also show Ks of exon core domains and pseudoESE, i.e. hexamers of the same underlying nucleotide content as ESEs but not necessarily identified as being functional ESE. We consider 20 intron size bins apportioned so that all bins contain the same number of exon ends for concatenation, the numbers given reflecting the upper intron size limit of each bin.

Fig.1.A.

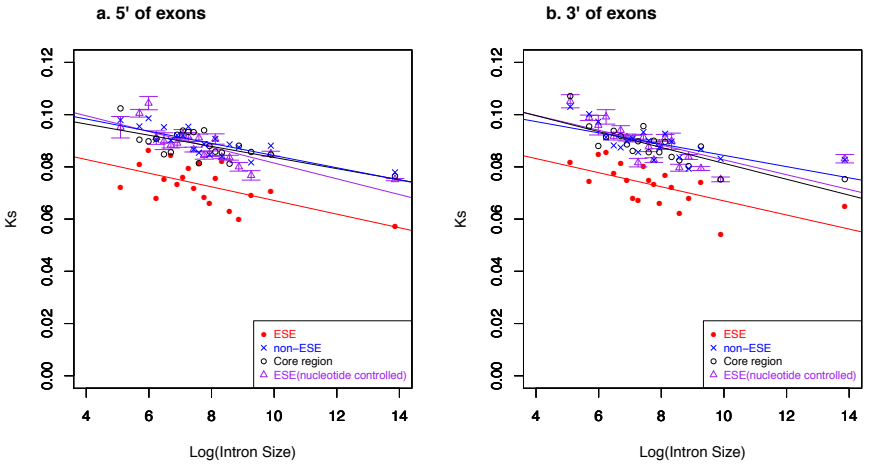

Fig.1.B.

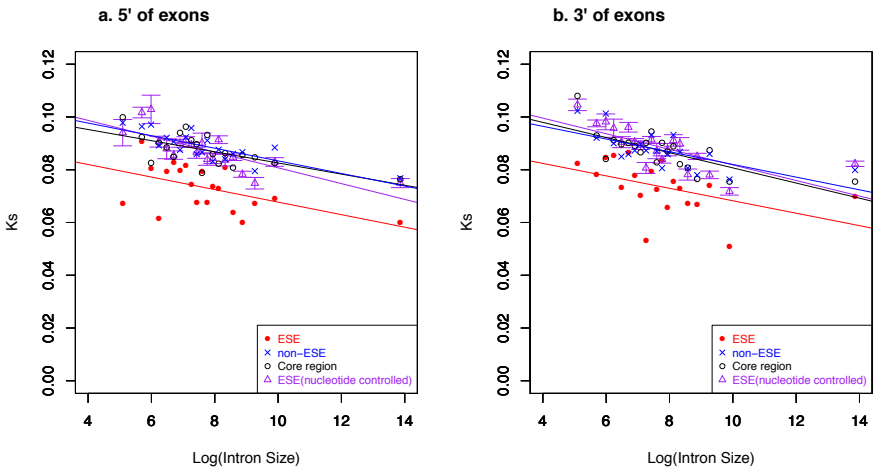

Fig.2. The degree of selective constraint on ESE sequences at exon ends as a function of the Log of flanking intron size for two ESE datasets (Fig.2.A: INT3, Fig.2.B: INT3_400). For definition of constraint see main text. For intron size definition see Fig 1. Note that in all cases constraint appears stronger when intron sizes are larger, although using 20 bins the trends are not significant.

Fig.2.A.

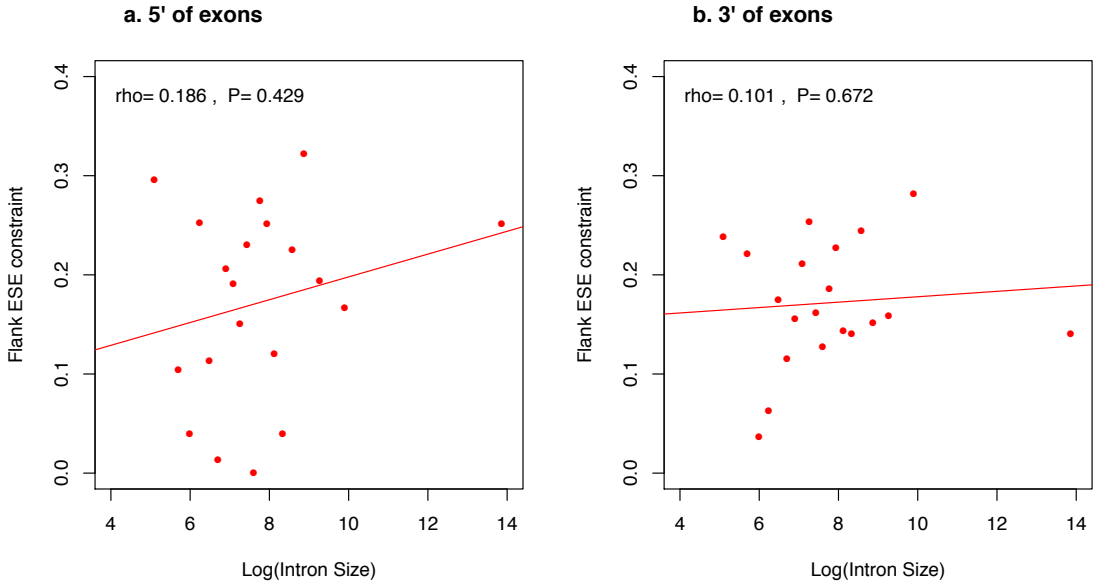

Fig.2.B.